AD ________________

Award Number: DAMD17-94-J-4371

TITLE: Computer Aided Breast Cancer Diagnosis

PRINCIPAL INVESTIGATOR: Carey E. Floyd, Ph.D.

CONTRACTING ORGANIZATION: Duke University Medical Center
Durham, North Carolina 27710

REPORT DATE: October 2000

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	October 2000	Final (23 Sep 94 - 22 Sep 00)

4. TITLE AND SUBTITLE
Computer Aided Breast Cancer Diagnosis

5. FUNDING NUMBERS
DAMD17-94-J-4371

6. AUTHOR(S)
Carey E. Floyd, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Duke Universitv Medical Center
Durham, North Carolina 27710

E-Mail: cef@deckard.mc.duke.edu

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
This report contains colored photos

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT ***(Maximum 200 Words)***

The long term goal of this work is to develop a computer aid for the decision for breast biopsy. In this project, an artificial neural network was developed to predict the outcome of biopsy from the mammographic features described in the BIRADS lexicon. The focus has been to optimize the performance of the artificial neural network as well as to investigate alternate decision models. Cases were acquired from Duke University, and all cases included biopsy proof of the presence or absence of malignancy. In testing, the computer aid was found to be able to reduce the number of benign lesions that would be biopsied by 40% while missing 2% of the malignancies. Several alternative classifier designs were found to be offer promise as well. The deployment of this system into regional care facilities and into private mammography practices could facilitate transferring the expertise currently present in only a few tertiary care centers to the public at large and to smaller and more rural settings and thus improve access for under-served populations.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Breast Cancer	27
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the U.S. Army.

✓ Where copyrighted material is quoted, permission has been obtained to use such material.

NA Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

✓ Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

NA In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and use of Laboratory Animals of the Institute of Laboratory Resources, national Research Council (NIH Publication No. 86-23, Revised 1985).

✓ For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal law 45 CFR 46.

NA In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institute of Health.

NA In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA molecules.

NA In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.

Carey [illegible] 24 Jan 2001

PI – Signature Date

Table of Contents

Front Cover	1
Report Documentation Page (SF 298)	2
Foreword	3
Table of Contents	4
Introduction	5
Report on Research for 2000	5
Conclusion (Discussion)	26
References	27

Report of the Progress on Grant DAMD17-94-J-4371

For the Period 23 September 1994 to 23 October 2000

Introduction

Biopsy is considered to be the definitive test to rule out breast cancer for those patients who participate in breast screening examinations and whose mammograms are interpreted as having suspicious findings. Excisional biopsy is a sensitive and specific test for breast cancer[1]. If the cost of excisional biopsy were minimal, this would be an ideal test for breast cancer malignancy. Unfortunately, the cost of this procedure in both monetary and emotional terms, is significant [2,3,4]. Unfortunately, to achieve a high sensitivity for detecting cancer, many women with mammographic findings due to benign processes undergo biopsy. The false positive rate for the decision to biopsy is currently between 66% and 90%. The goal of the work described here is to design a decision tool to support the decision to biopsy. This decision aid must maintain the current high detection rate for true cancers while accurately ruling out some of the benign cases and thus avoiding unnecessary biopsies.

The problem of classifying suspicious mammographic lesions as benign or malignant is recognized as a difficult practice. There is considerable variation in the skill with which the task is achieved even within the group of radiologists who specialize on this task. The radiographic manifestation of breast cancer is not well enough understood from a fundamental scientific basis to allow an accurate theoretical predictive model to be

constructed from first principals. There is no accurate deterministic model for relating mammographic findings to biopsy outcomes although some general rules are accepted. Examples of these rules are "Older women are more likely to develop breast cancer than young women." "If the margin of a mass appears well circumscribed, the mass is likely to be benign." Unfortunately, the sensitivity and specificity of those rules that are generally agreed upon is not sufficient to allow a strict implementation over the full range of cases that are encountered in clinical practice. While rule based expert system s have enjoyed success in some medical diagnostic tasks, and there are expert mammographers whose diagnostic performance would qualify them as experts, the construction of rule-ased expert systems for this diagnostic task has met with limited success. This is quite possibly due to the difficulty of describing the logical and analytic process used by the experts in a form that can be used by other mammographers. Atypical difficulty with the expert system approach is the description and encoding of the input data: two radiologists often will use similar, but not exactly the same descriptions for a given lesion. It is often difficult to overcome instability in a model due to this potential ambiguity in the input data. Another difficulty for strict rule-based systems is that the descriptors used as inputs to the model can be nonspecific: two lesions with similar descriptions can have opposite outcomes at biopsy. These arguments indicate that an example-based technique would be more appropriate. This is supported by realizing that radiologists are trained by repeatedly examining sample cases with known outcomes that are maintained in a medical school's teaching files. The focus of this research has been in developing and evaluating data driven models, specifically artificial neural networks (ANN) for the task of predicting the outcome of biopsy given the description of mammographic lesions as

inputs. This work has been facilitated by the growing acceptance of BI-RADS as a standardized lexicon for mammographic case reporting.

Progress in this project is demonstrated through the 47 publications supported in part by this grant.

The publications included 13 peer-reviewed journal, 19 manuscripts in conference proceedings, and 15 conference abstracts of presentations

1. Floyd CE Jr, Yun AJ, Sullivan D, Kornguth P. Prediction of Breast Cancer Malignancy using an Artificial Neural Network. *Cancer* 74:2944-2948,1994.

2. Baker JA, Kornguth PJ, Lo JY, Williford ME, Floyd CE Jr. Breast Cancer: Prediction with Artificial Neural Network Based on BIRADS Standardized Lexicon *Radiology* 196;817-822; 1995.

3. Lo JY, Baker JA, Kornguth PJ, Floyd CE Jr. Computer-aided diagnosis of breast cancer: artificial neural network approach for optimized merging of mammographic features. *Academic Radiology*, 2;841-850;1995.

4. Baker JA, Kornguth PJ, Lo JY, Floyd CE Jr. Artificial Neural Network: Improving the Quality of Breast Biopsy Recommendations. *Radiology*;198;131-135; 1996.

5. Baker JA, Kornguth PK, Floyd CE Jr. Breast Imaging Reporting and Data System Standardized Mammography Lexicon: Observer Variability in Lesion Description *Amer. J. Roent.*;166;773-778; 1996.

6. Lo JY, Baker JA, Kornguth PJ, Igelhart R, Floyd CE Jr. Predicting Breast Cancer Invasion From BI-RADS Mammographic Features Using Artificial Neural Networks On The Basis Of Mammographic Features. *Radiology*, 203;159-163; 1997.

7. Tourassi GD, Floyd CE Jr. The Effect of Data Sampling on the Performance Evaluation of Artificial Neural Networks in Medical Diagnosis. *Medical Decision Making*; 17;186-192; 1997.

8. Lo JY, Baker JA, Kornguth PJ, Floyd CE Jr. Effect of Patient History Data on the Prediction of Breast Cancer from Mammographic Findings with Artificial Neural Networks. *Acad Radiol*, 6;10-15; 1999.

9. Gavrielides MA, Lo J, Vargas-Voracek R, Floyd CE Jr. Segmentation of suspicious clustered microcalcification in mammograms. *MedicalPhysics*; 27(1):13-22; 2000.

10. Floyd C.E., Jr.,Lo J.Y., Tourassi G.D., Breast Biopsy: Case-Based Reasoning Computer-Aid Using Mammography Findings for the Decision to Biopsy, In press to American Journal of Roentgenology (AJR) 2000.

11. Floyd C.E., Jr.,Lo J.Y., Tourassi G.D., Breast Biopsy: Case-Based Reasoning Computer-Aid Using Mammography Findings for the Decision to Biopsy, American Journal of Roentgenology (AJR) 175:1-6, 2000.

12. Markey M.K, Lo J.Y., Vargas-Voracek R., Tourassi G.D., Floyd C.E.Jr., "Perceptron Error Surface Analysis: A Case Study in Breast Cancer Diagnosis", submitted to IEEE Transactions in Medical Imaging

13. Lo JY, Markey MK, Baker JA, and Floyd CE, Jr, "Cross-institution evaluation of BI-RADS model for mammographic diagnosis of breast cancer," submitted, (2001).

Conference Proceedings:

1. Floyd CE Jr, Yun A, Lo JY, Tourassi GD, Sullivan D, Kornguth P. Prediction of Breast Cancer Malignancy for Difficult Cases using and Artificial Neural Network. In *World Congress on Neural Networks,* International Neural Network Society Annual Meeting (INNS), 1:I127-I132, 1994.

2. Floyd CE Jr, Grissom A, Yun J, Lo JY, Dovan M, Humphrey L, Sullivan DC, Kornguth PJ. Computer-Aided Breast Cancer Prediction: Integration of a Mammography Findings Database with an Artificial Neural Network. In *Computer Applications to Assist Radiology, Symposium for Computer Assisted Radiology (SCAR),* 255-260, 1994.

3. Floyd CE Jr, Soo MS, Tourassi GD, Kornguth PJ. Computer aided prediction of breast implant rupture based on mammographic findings. In *Proceedings of the International Society for Optical Engineering (SPIE),* 2434; 471-477, 1995.

4. Lo JY, Grissom AT, Floyd CE Jr, Kornguth PJ. Computer-aided diagnosis of mammograms using an artificial neural network: Merging of standardized input features from the ACR lexicon. In *Proceedings of the International Society for Optical Engineering (SPIE),* 2434;571-578; 1995.

5. CE Jr, Lo JY, Tourassi GD, Kornguth P. Computer aided diagnosis using genetic algorithms and neural networks. In *World Congress on Neural Networks,* International Neural Network Society Annual Meeting (INNS), II-863-866;1995.

6. Tourassi GD, Floyd CE Jr. Sostman HD, Coleman RE. Performance evaluation of an artificial neural network for the diagnosis of pulmonary embolism using the cross-validation, jackknife, and bootstrap methods: a comparison study. In *World Congress on Neural Networks,* International Neural Network Society Annual Meeting (INNS),II-897-900;1995.

7. Lo JY, Baker JA, Kornguth PJ, Floyd CE Jr. Computer-aided diagnosis of mammography: Artificial neural networks for optimized merging of standardized BIRADS features. In *World Congress on Neural Networks,* International Neural Network Society Annual Meeting (INNS), II-885-888, 1995.

8. CE Jr, Use of genetic algorithms for computer-aided diagnosis of breast cancer from image features. In *Proceedings of the International Society for Optical Engineering (SPIE),* 2710;51-58;1996.

9. Lo JY, Floyd CE Jr, Kornguth PJ. Computer-aided diagnosis of mammography using an artificial neural network: predicting the invasiveness of breast cancers from image features. In *Proceedings of the International Society for Optical Engineering (SPIE),* 2710: 725-732; 1996.

10. Lo JY, and Floyd CE, Jr, "Analysis of error surfaces of neural network applied to computer-aided diagnosis in mammography," World Congress on Neural Networks '96 (International Neural Network Society 1996 Annual Meeting), Lawrence Erlbaum Associates, Inc., San Diego, CA, 1240 (1996).

11. Lo, J.Y. and Floyd CE, Jr, "Self-organizing maps for analyzing mammographic findings," 4, Karayiannis NB, Ed., IEEE International Conference on Neural Networks, IEEE, Houston, TX, 4: 2472-4 (1997).

12. Vargas-Voracek R, Floyd CE Jr. Computer-Aided Diagnosis for Early Detection of Breast Cancer from Mammograms. Susan G. Komen Breast Cancer Foundation "Reaching for the Cure" National Grant Conference. (1998).

13. Vargas-Voracek R, Floyd CE Jr. Markov-Random Field Texture Model for Automatic Breast Parenchyma Characterization. Accepted for presentation at the 84th Radiological Society of North America (RSNA) Scientific Assembly and Annual Meeting. November 29-December 4, 1998.

14. Lo JY, and Floyd CE, Jr, "Computer-aided diagnosis of breast cancer," Doi K et al., Ed., First International Workshop on Computer-Aided Diagnosis, Elsevier Science, Univ. of Chicago, Chicago, IL, 1182 (ICS 1182): 221-5 (1998).

15. Floyd CE, Jr, Lo JY, and Baker JA, "Prediction of breast biopsy outcomes from mammographic findings," Doi K et al., Ed., First International Workshop on Computer-Aided Diagnosis, Elsevier Science, Univ. of Chicago, Chicago, IL, 1182 (ICS 1182): 193-200 (1998).

16. Floyd CE, Jr, Lo JY, Tourassi, GD, "Case-Based Reasoning as a Computer Aid to Diagnosis," Medical Imaging 1999: Image Processing, Hanson KM, Ed., *Proc. SPIE*, 3661:486-489, 1999.

17. Vargas-Voracek R, Floyd CE Jr. Hierarchical Markov-Random Field Texture Modeling for Mammographic Structure Segmentation Using Multiple Spatial and Intensity Image Resolutions. 1999 Medical Imaging Symposium. International Society for Optical Engineering (SPIE). February 20-26, 1999.

18. Tourassi GD, Floyd CE Jr, Lo JY. A Constraint Satisfaction Neural Network for Medical Diagnosis. 1999 International Conference on Neural Networks (ICNN), Washington, DC.

19. Tourassi GD, Floyd CE, Jr, and Lo JY, "Use of constraint satisfaction neural network for breast cancer diagnosis and dynamic scenarios simulation," Medical Imaging 2000: Image Processing, Hanson KM, Ed., SPIE Medical Imaging 2000: Image Processing, Proc. SPIE 3979: 46-54 (2000).

Presentations and Abstracts:

1. Lo JY, Floyd CE Jr, Tourassi GD. Artificial Neural Network for Diagnosis in Radiology. Lo SCB, Ed., Computer-Aided Diagnosis Workshop, Georgetown University Medical Center, Washington, DC. 22, 1994
2. Floyd CE Jr, Lo JY, Baker JA, Kornguth PJ: Interactive Computer-Aided Diagnosis of Breast Cancer. *Radiology* 197P:533, 1995.
3. Baker JA, Kornguth PJ, Floyd CE Jr: Interobserver variability in Radiologist's Use of the BI-RADS Mammography Lexicon. *Radiology* 197P:242, 1995.
4. Baker JA, Kornguth PJ, Lo JY, Floyd CE Jr: Artificial Neural Network for the Prediction of Breast Cancer with the BI-RADS Standardized Lexicon. *Radiology* 197P:242, 1995.
5. Lo JY, Baker JA, Kornguth PJ, Floyd CE Jr: Application of Artificial Neural Networks to the interpretation of Mammograms on the Basis of the Radiologist's Impression and Optimized Image Features. *Radiology* 197P:242, 1995.
6. Lo JY, Baydush AH, Baker JA, Kornguth PJ, Floyd CE Jr: Computer-Aided Diagnosis of Breast Mass Malignancy with Automated Feature Extraction and Artificial Neural Networks. *Radiology* 197P:425, 1995.
7. Lo JY, Baker JA, Floyd CE, Jr, "Artificial neural networks for the prediction of breast cancer invasiveness by using Breast Imaging and Reporting Data System mammography lexicon," Radiology 201P, 370; 1996.
8. Lo JY, and Floyd CE Jr. Analysis of Error Surfaces of Neural Network Applied to Computer-Aided Diagnosis in Mammography. World Congress on Neural Networks '96 (International Neural Network Society 1996 Annual Meeting), Lawrence Erlbaum Associates, Inc., San Diego, CA 1240, 1996.
9. Lo JY, Baker JA, Frederick ED, Kornguth PJ,and Floyd CE Jr. Predicting breast lesion malignancy and invasion using the BI-RADS mammography lexicon, Radiology 205(P), 447 (1997).
10. Floyd CE, Jr, Lo JY, Baker JA, Frederick ED, Kornguth PJ: "Computer Aid for the Decision to Biopsy" Radiology 205P, 740; 1997.
11. Lo JY, Baker JA, Frederick ED, Kornguth PJ, Floyd CE, Jr: "Predicting Breast Lesion Malignancy and Invasion Using the BI-RADS Mammography Lexicon" Radiology 205P, 447; 1997.
12. Floyd, CE, Jr: "Computer Aided Diagnosis in Medical Imaging at Duke University," Presented at the AAPM Annual Meeting, San Antonio Texas 1998.
13. Baker JA, Frederick ED, Lo JY, Kornguth PJ, and Floyd CE Jr. Incorporation of an Artificial Neural Network into Clinical Mammography to Reduce Benign Breast Biopsies. *AJR Supplement* 170:84, 1998.
14. Lo JY, Kornguth PJ, Floyd CE Jr. Multi-Institution Evaluation of BIRADS Breast Cancer Prediction Model. *Radiology* 209(P):271, 1998.

15. Floyd CE Jr., Lo JY, Baker JA, Kornguth PJ Multi-Institution Evaluation of Case-Based Reasoning for Breast Cancer Prediction. *Radiolog* 213(P), 334 1999

ANN systems for the prediction task have been constructed and evaluated using a growing database of mammographic cases that were sent to biopsy with the results known. This work has been successful and has resulted in xxx peer-reviewed publications, xxx invited presentations, xxx competitive presentations, and has served to foster further research efforts for constructing decision aids for the diagnosis of breast cancer as demonstrated by the 11 funded grants that have been awarded since the start of this project. The ideas that developed into these other grants were realized from this work.

In the latter part of this research grant, the investigative emphasis has shifted from algorithm development to clinical evaluation reflecting the shift from specific aim one to specific aim two. After conducting several preliminary sessions with mammographers using the system including three years of presentation of a live computer demonstration version at the Radiological Society of North America InfoRad exhibit, several important question shave been identified regarding the user interface of the system. The first is the question of how the results should be presented to the mamographer. An informal exit interview with mammographers who used the system indicated that 70% preferred a probabilistic output where the mammographer would be given a number between 0 and 100 to indicate the estimated percent probability that the case in question was malignant. The other 30% of he users did not want a probability, they wanted a hard decision to biopsy or not to biopsy. For these clinicians, the system threshold would be set to some value and the binary result would be presented. All users, especially those preferring the hardwired decision threshold, desired some indication of the certainty with which the

decision was presented. Several individuals expressed an interest in being presented with "similar" cases from which the neural network was trained. These reasonable requests initiated the ideas that lead to the development of the CBR. In an effort to provide similar "example cases", it was realized that cases with similar findings would generate similar ANN outputs even though these would not provide a complete or unique subset. A case findings matching algorithm was implemented using a relational database (MicroSoft ACCESS ™) to simplify and speed the coding. It was later found that this implementation also dramatically improved the speed of execution. With this case matching tool, different definitions of what constituted a similar case could be investigated. It was found that depending on how strict the definition of similarity, the existing database could provide between 10 and 100 similar cases for each new case to be evaluated. While beyond the scope of this investigation, given these cases identifications, it would be straightforward to present digital version of the cases on a monitor to provide partial explanation of the ANN result. It was natural to compute the fraction of malignancies to total cases within the set of matched cases. With the use of this fraction as a decision variable, a predictive tool was naturally implemented. While the evolution of this technique proceeded as described above from an effort to provide explanation to the mammographer for the recommendation suggested by the ANN, it was soon recognized that the resulting algorithm was an instantiation of a simple case-based reasoning system.

A preliminary investigation was performed to better understand the relationship of findings to malignancy within the framework of the BI-RADS reporting lexicon. A Case-

Based Reasoning (CBR) approach was selected for this study since we wished to examine the cases and the similarity between them. In this context, a CBR was developed and evaluated by its ability to predict the outcome of biopsy from mammographic findings reported in the BI-RADS lexicon. To classify a given test case as benign or malignant, CBR was implemented by comparing the case to all previous cases, selecting those cases with were similar with regards to their findings and examining the outcomes for those similar cases. A decision variable was formed as the "malignancy ratio: computed as the ratio of the number of malignant cases to the total number of similar or "matched" cases. Performance was evaluated by generating an ROC curve from the true positive fraction and the false positive fraction as the threshold was applied to the malignancy ratio.

The system is implemented as follows. The mammograms are read by clinicians using a standard reporting lexicon (BI-RADS™). These findings are compared to a database of findings from cases with known outcomes (from biopsy). The fraction of similar cases that were malignant is returned. The clinician can then consider this result when making the decision regarding biopsy. . This malignancy fraction is an intuitive measure that can be readily included in the medical decision. This approach is intuitive. The CBR answers the question "Of all cases that are similar to this one, how many were malignant at biopsy?" This process is similar to that followed by the clinician when considering the same problem.

Methods

The CBR was implemented as a case retrieval engine in a relational database framework. In this context, it functions as a query of a table of cases and outcomes. To predict the outcome for a new test case, the test case is compared to each case in the database through a matching rule. The prediction is the ratio of the number of malignant to the total number of cases that match.

The components of the system include the case encoding and the matching rule. The cases are encoded through a subset of the categorical BI-RADS^{TM} image findings and the patients' age. For the initial experiment, similarity is defined as an exact match of some subset of the findings. The database has been described previously[5] and was restricted for this feasibility study to the first 500 cases since the properties of this set were well understood and numerous previous studies had been conducted on this set.. Of these 500, 232 of the cases described masses, 192 cases described microcalcifications, and 29 cases described masses and microcalcifications associated that were associated with the mass. The remaining 47 cases did not describe either a mass or a calcification but were architectural distortions, asymmetric breast density, focal asymmetric density, and/or asymmetric breast tissue. Malignant outcomes were reported for 174 (35%) of the biopsies while 326 (65%) were found to be benign resulting in a Positive Predictive Value (PPV) of 35%.

The input features were restricted to those that had been found to have the highest independent predictive power in our earlier studies and included the patient age, and , for masses, the mass margin, mass size, mass density, and mass shape, while for

calcifications included calcification description, calcification number, calcification distribution, and special cases/associated findings.

Performance of the predictive system was evaluated through a round-robin technique in which: a test case is selected from the dataset, the database is formed from all of the other remaining cases, all of these remaining cases are compared to the test case and those that match are selected. The malignancy fraction is found for the set of matching cases. The testing example is replaced in the set and another is removed, the resulting system is evaluated and this is repeated until all examples have been used as testing cases. A threshold is applied to the set of malignancy fractions and the sensitivity and false positive fraction are plotted as the threshold is applied at each value of the malignancy fraction. A Receiver Operating Characteristic ROC curve is plotted from these ordered pairs. The ROC areas were computed from the resulting curve using Newton's method of integration. Used as a performance measure, ROC area gives equal significance to the sensitivity and specificity resulting from the application of a specific threshold. It is clear that sensitivity has higher priority than specificity for this problem since, while there is a need to reduce the number of benign biopsies, there is a greater cost incurred by missing a malignancy than by performing a biopsy on a benign lesion. As a consequence, a more appropriate measure could be formed by concentrating on the performance at high sensitivity. To concentrate on this region, three other measures are presented: the partial ROC area reported for sensitivity greater than 90% , and , the specificity at sensitivities of 98% and 100%.

Results

Performance for the ANN is presented in table 1 and described below. From a previous study on the predictive power of the findings using linear discrimanent analysis (LDA) the following six findings were found to contribute significantly: Age, Mass Margin, Mass Density, Calcification Description, Calcification Distribution, and Associated Findings. Requiring an exact match on all six features resulted in an ROC area of 0.77 but with very poor (<1%) specificity at high sensitivities of 90% and higher.

Table 1

Performance of Case Based Reasoning				
Matching Rule	ROC Az	Partial ROC Az	Specificity at 100% Sensitivity	Specificity at 98% Sensitivity
6 findings	0.77	0.016	<0.01	0.012

Table 1 CBR performance.

The histogram for the malignant (positive) and benign (negative) cases as a function of malignancy fraction is shown in fig. 1. The striped boxes indicate the negative cases while the solid boxes show the positive cases.

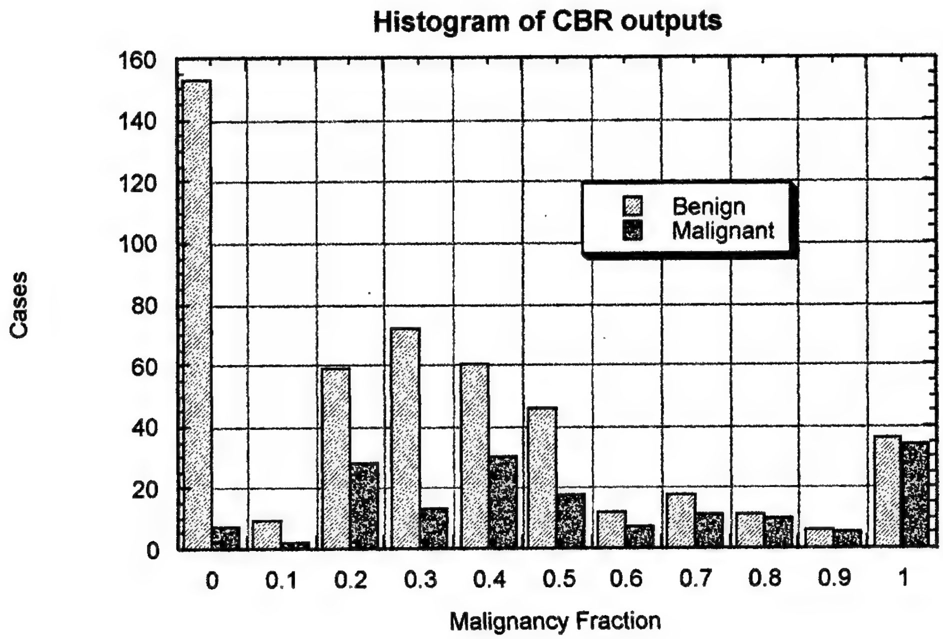


Fig. 1 Histogram of CBR outputs.

The ROC curve is shown in fig.2.

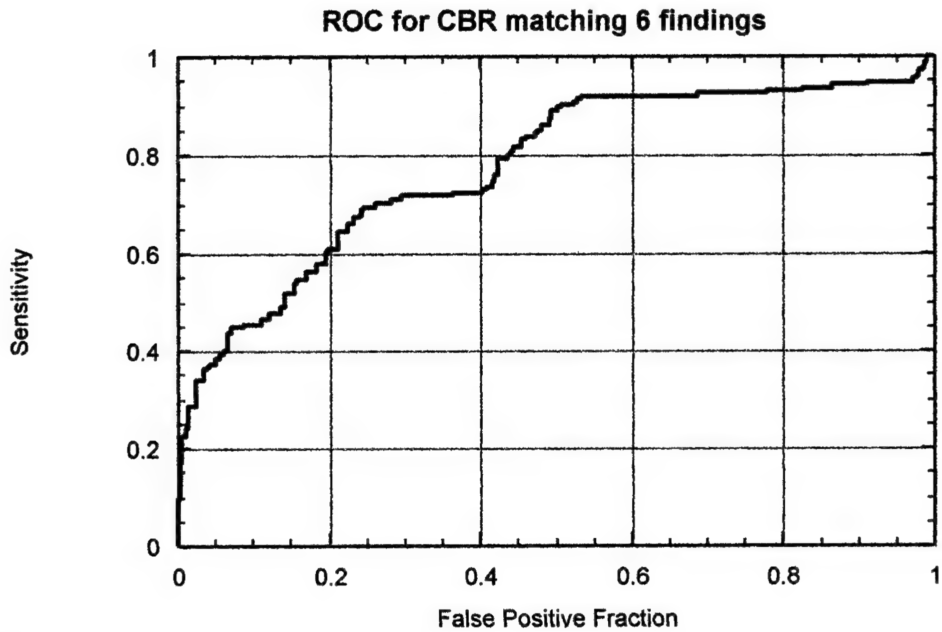


Fig. 2 ROC curve for CBR with exact match of 6 findings and age within 5 years.

Less than 0.12 seconds are required to predict the malignancy ratio for a new case with the system running in a non-optimized ACCESS™ (Microsoft Inc., Redmond Washington) database language on a Pentium II 300Mhz personal computer.

To date (September 2000), we have compiled a database of 1300 cases that were examined at diagnostic mammography and were referred to biopsy at Duke University Medical Center. Histograms for the most significant finding for mass cases (Mass Margin) is shown in fig. 3 for both benign and malignant outcomes.

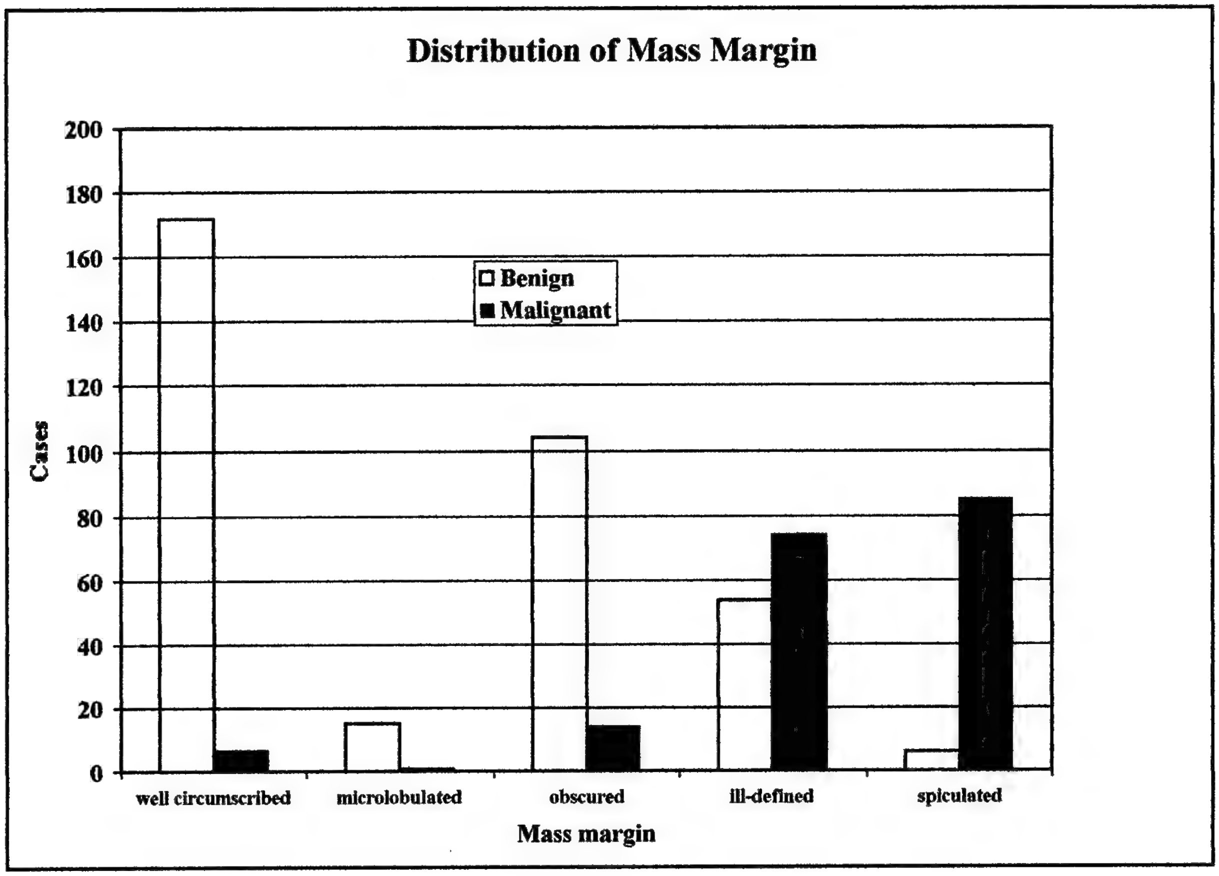


Fig 3 Distribution of the Mass Margin feature over 532 mass cases. The distribution for benign cases is shown as the open bars while the distribution for malignant cases is shown as the solid bars.

Histograms for the most significant finding for calcification cases (Calcification Description) is shown in fig 4. for both benign and malignant outcomes.

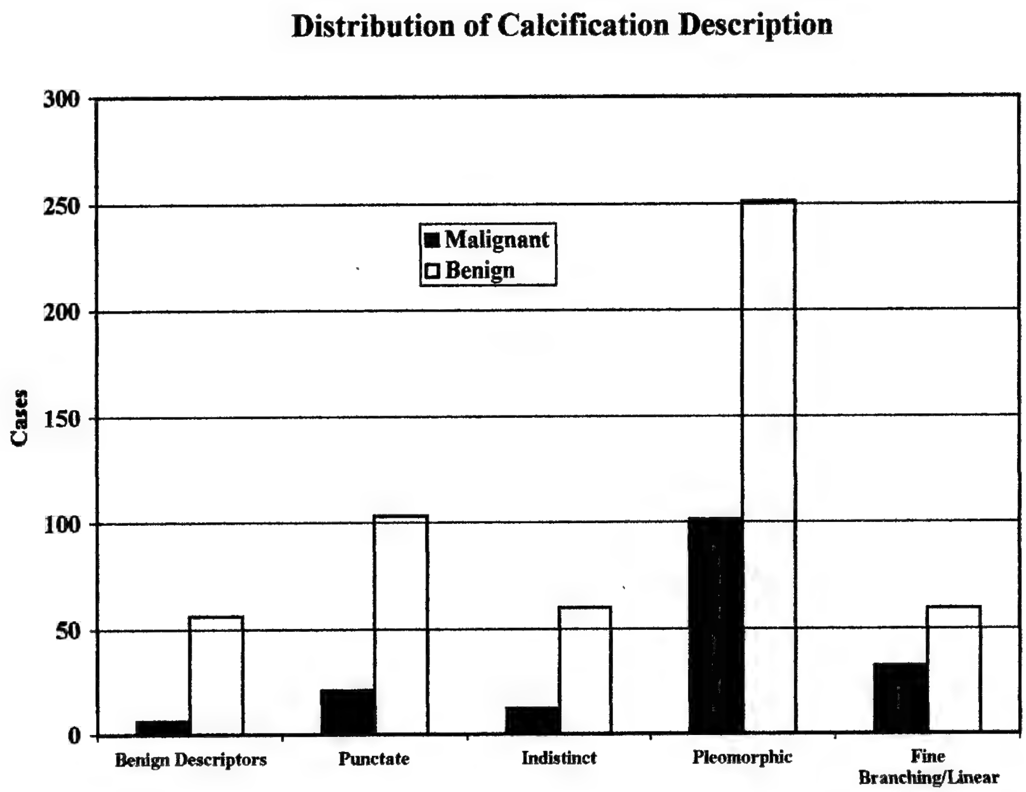


Fig 4. Distribution of the Calcification Description feature over the 529 calcification cases. The distribution for cases with benign outcomes is shown as the open bars while the distribution for cases with malignant outcomes is shown as the solid bars. Following the second edition of BIRADS, the plot has combined the "benign" calcification descriptions into one category which now includes the descriptions of milk of calcium-like, rim, skin, vascular, spherical, suture, coarse, large rod-like, round, and dystrophic.

The potential for predictive power is obvious for these two findings. A complete analysis of these data is in progress and will be submitted for publication in 2001.

Subgroups of cases

For 1270 cases in the existing database (30 suspended cases are currently in final review), the population of major sub-sets are shown in table 2 along with the fraction of cases in the sub-set that were malignant. The change from previous field currently is being re-structured and was unavailable for this table.

Population of subgroups defined by lesion type.

Class	cases	% of total	%malignant (PPV)
All	1270	100	34
Mass Only	532	42	34
Calc Only	529	42	33
No Mass No Calc	147	12	35
Mass and Calc	62	4	54

Table 2: showing the population of major subgroups defined by lesion type.

Only the class with both mass and calcification features has an obvious difference in the fraction of malignancy.

Subgroup by unsupervised clustering algorithms

AUTOCLASS

In preliminary investigations, between 3 and 5 clusters were typically formed using the program AUTOCLASS. An artificial neural network (ANN) was developed and evaluated on the two common types of breast lesions, masses and calcification clusters. When the ANN was trained on a data set containing both masses and calcifications, the area (Az) under the receiver operating curve (ROC) for all the cases was 0.86 ± 0.02 and the partial area index (PAI) for TPF0 > 0.9 was 0.50 ± 0.05. The performance of the ANN on cases containing only masses ($Az = 0.95 \pm 0.01$, $PAI = 0.77 \pm 0.05$) was quite different from the performance on cases containing only calcifications ($Az = 0.70 \pm 0.04$, $PAI = 0.29 \pm 0.06$). A logistic regression and radiologists' gut assessment both exhibit a similar difference in performance on masses versus calcifications. This indicates a clear motivation for improvement for the cases with calcifications.

3.3.3 Constraint-Satisfaction

A constraint satisfaction neural network has been constructed and preliminary evaluations were performed on the first 500 cases in the database. As table 3 shows, the CSNN provides competitive performance as a classifier.

Classifier	ROC Area Index (A_Z)	SPECIFICITY at 95% Sensitivity	PPV
CSNN	0.84 ± 0.02	50%	50%
BP-ANN	0.87 ± 0.02	52%	51%
Mammographers	0.82 ± 0.02	37%	45%

Table 3: CSNN diagnostic performance when applied as a classifier. Previously published performance of experienced mammographers and a backpropagation artificial neural network (BP-ANN) are included for comparison purposes.

Case-Based Reasoning

In preliminary studies, we constructed a simple CBR system to classify cases referred for biopsy. The CBR was evaluated on a set of 500 cases. A receiver operating characteristic curve for the CBR performance is shown in fig. 5 below. Note the encouraging behavior at high sensitivity. The sensitivity remains very high as the false positive fraction (FPF) decreases and does not significantly decrease until the FPF has dropped to 0.6 (specificity of 0.4). With a threshold of 0.2, 126 benign biopsies could be avoided at a cost of 2 missed malignancies.

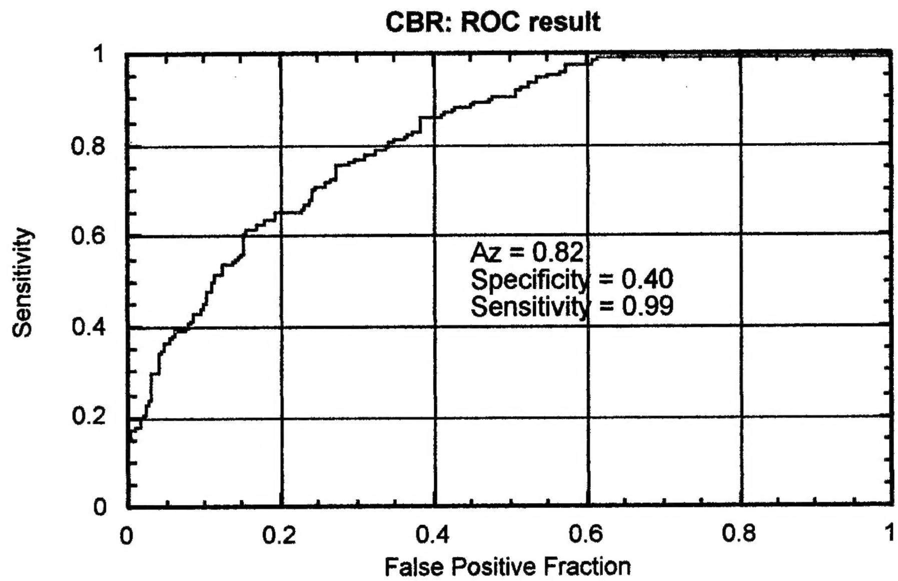


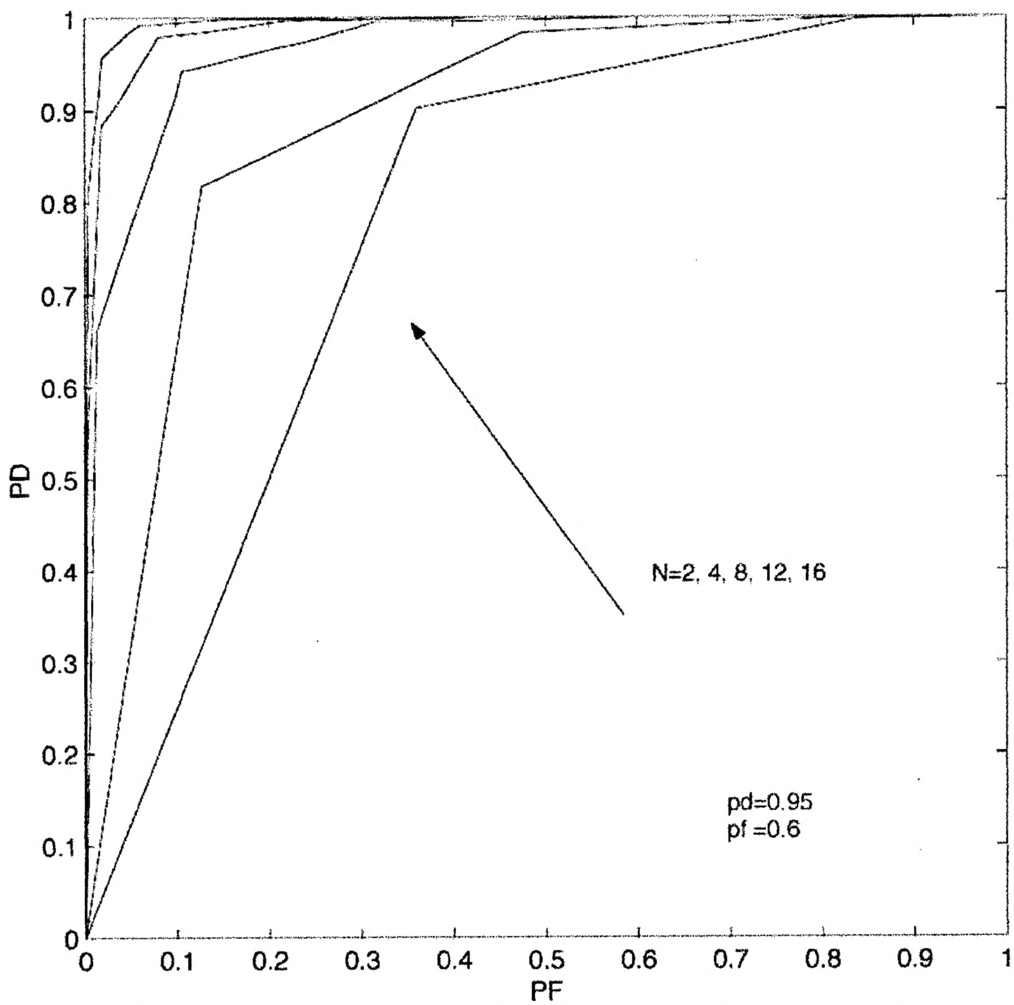
Fig. 5. ROC plot of CBR output values for all benign and malignant cases.

The portion of the ROC curve that is of greatest interest is the region of greatest true-positive fraction (i.e. highest sensitivity) since few radiologists or patients would be willing to under diagnose breast cancer for the sake of high specificity. At sensitivity of 0.98 (relative to all biopsied lesions) the specificity of some of our previous classifiers has been as high as 0.4. Thus, almost 40% the benign biopsies could have been avoided at the cost of missing 2% of the malignancies. The positive predictive value would be increased from 35% to 46%. This study shows that classifiers using the BI-RADS^{TM} lexicon as inputs has the potential to improve the positive predictive value of the

recommendation for breast biopsy. The ROC curve was plotted from the data directly, not from a fit to the data.

3.2.4. Combined local classifiers

The optimal signal processing technique allows for a theoretical analysis for an upper bound on the performance of the proposed combination of local "experts" or classifiers.



ROC Performance Bounds of Optimal Combination of Decisions of N Experts

Fig. 6 illustrates the upper bound ROC for optimally combining the outputs of N experts, if each expert is operating at a sensitivity(TPF) of 0.95 and a false positive fraction (FPF) of 0.60.

Fig.6 represents an upper bound on the improvement which would be realized if all of the experts are independent. In practice, the local classifiers (experts) are likely to be correlated to some degree and so the actual improvement is expected to be less. While the performance gained through this technique does depend on the independence of the experts, the validity of the technique does not. The encouraging aspect of the ROC curves in this figure is the dramatic decrease in FPF at high sensitivities as N, the number of combined experts, is increased. The ROC curves in this figure were computed assuming the same fixed operating point for each expert, but the technique is easily applied to combine experts with different operating points. In addition, individual operating points can be optimized if continuous outputs are available as they will be for most of the local classifiers proposed.

Discussion

The work described here has resulted in an ANN with performance that can significantly improve the current poor specificity of the clinical referral to breast biopsy without sacrificing the sensitivity. Several option s have been investigated for alternate decision strategies and several show promising performance. The work performed under this grant has served to generate seed ideas that will be pursued further. The extensive publications resulting from this work form a solid framework with which to pursue this further work.

References

1. Kopans DB. The positive predictive value of mammography. *AJR. American Journal of Roentgenology* **1992**; 158: 521-526
2. Dixon JM and John TG. Morbidity after breast biopsy for benign disease in a screened population. *Lancet* **1992**; 1: 128
3. Helvie MA, Ikeda DM, and Adler DD. Localization and needle aspiration of breast lesions: complications in 370 cases. *AJR. American Journal of Roentgenology* **1991**; 157: 711-714
4. Schwartz GF, Carter DL, Conant EF, Gannon FH, Finkel GC, and Feig SA. Mammographically detected breast cancer: nonpalpable is not a synonym for inconsequential. *Cancer* **1994**; 73: 1660-1665
5. Lo JY, Baker JA, Kornguth PJ, and Floyd CE, Jr. Effect of patient history findings on predicting breast cancer from mammograms using artificial neural networks. *Academic Radiology* **1998**; 7.